DOCUMENT RESUME

ED 050 143 TM 000 530

AUTHOR	McRae, Douglas J.
TITLE	Multivariate Cluster Analysis.
INSTITUTION	McGraw-Hill Book Co., Monterey, Calif.
PUB DATE	Feb 71
NOTE	22p.; Paper presented at the Annual Meeting of the American Educational Research Association, New York, New York, February 1971
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Ability Grouping, *Algorithms, Cluster Grouping, Criterion Referenced Tests, Data Analysis, Educational Research, Effective Teaching, *Elementary School Students, Grades (Scholastic), *Grouping (Instructional Purposes), Homogeneous Grouping, Mathematics, *Statistical Analysis

ABSTRACT

Procedures for grouping students into homogeneous subsets have long interested educational researchers. The research reported in this paper is an investigation of a set of objective grouping procedures based on multivariate analysis considerations. Four multivariate functions that might serve as criteria for adequate grouping are given and discussed; a method for optimizing these functions is also described. The set of procedures is illustrated through application to data from two samples of students, each student with scores on either ten or eleven subtests of a criterion referenced mathematics inventory. The results indicate that the procedures discussed provide a promising means for grouping students to minimize classroom heterogeneity. (Author)

Multivariate Cluster Analysis

Douglas J. McRae
CTB/McGraw-Hill
Del Monte Research Park
Monterey, California 93940

February, 1971

Paper presented at the 55th Annual Meeting of the American Educational Research Association, New York City, February 4 - 7, 1971.

MULTIVARIATE CLUSTER ANALYSIS

Douglas J. McRae
CTB/McGraw-Hill
Monterey, California

ABSTRACT

Procedures for grouping students into homogeneous subsets have long interested educational researchers. The research reported in this paper is an investigation of a set of objective grouping procedures based on multivariate analysis considerations. Four multivariate functions that might serve as criteria for adequate grouping are given and discussed; a method for optimizing these functions is also described. The set of procedures is illustrated through application to data from two samples of students, each student with scores on either ten or eleven subtests of a criterion referenced mathematics inventory. The results indicate that the procedures discussed provide a promising means for grouping students to minimize classroom heterogeneity.

The problem of grouping or clustering students into homogeneous subsets has been of interest to educational researchers for years. The rationale behind this interest is the assumption that teaching effectiveness is enhanced by homogeneity of ability, learning level, learning deficiencies, etc., in the students. The endpoint of this rationale is, of course, individualized instruction. This form of instruction, however, is either in the developmental state as far as educational hardware is concerned or, where hardware exists, is prohibitively expensive. Hence, educators still group or cluster students toward the aim of maximizing teacher effectiveness.

Traditionally, grouping procedures have been a subjective result of some objective measurement process. Student records in various subject areas are obtained from a variety of sources, for instance, previous grades, teacher evaluations, standardized tests; the administrator then sets a few basic decision rules and groups or clusters students on this basis. The efficiency of this procedure is open to question: Are the resultant groups in any sense maximally homogeneous? This paper discusses a set of objective procedures in which statistical and computer science technology is applied to the grouping procedure.

The line of thought followed for this work started with the suggestions of Sebestyen (1962). He suggested that one criterion for maximal grouping might be to minimize the sum of the distances from each observation to its group center. This is one of the criteria discussed below. Ball and Hall (1967) and MacQueen (1967) developed computer algorithms for optimizing this criterion; they also have investigated their procedures using real and artificial data. Friedman and Rubin (1967) extended the work by suggesting two new criteria based on multivariate analysis considerations; the Friedman and Rubin work included an algorithm for optimization

and empirical investigation. The technique as described in this paper includes four criteria for optimization and a combination of the MacQueen and Friedman-Rubin algorithms to accomplish the optimization.

Theory

This set of clustering procedures is motivated through consideration of an N x P data matrix, say $\underset{\sim}{X}$, where N refers to the number of students and P refers to the number of measurements available on each student. If one arbitrarily partitions the data matrix into g groups of students, then the cross-products matrix

$$\underset{\sim}{T} = \underset{\sim}{X}' \underset{\sim}{X}$$

may be partitioned into two matrices, $\underset{\sim}{W}$ and $\underset{\sim}{B}$, such that

$$\underset{\sim}{T} = \underset{\sim}{W} + \underset{\sim}{B}$$

where

$$\underset{\sim}{W} = \sum_{i=1}^{g} \sum_{j=1}^{n_i} (\underset{\sim}{X}_{ij} - \overline{\underset{\sim}{X}}_i)' (\underset{\sim}{X}_{ij} - \overline{\underset{\sim}{X}}_i) ,$$

and

$$\underset{\sim}{B} = \sum_{i=1}^{g} n_i \overline{\underset{\sim}{X}}_i ' \overline{\underset{\sim}{X}}_i ,$$

where g is the number of groups,

n_i is the number of students in the i^{th} group,

$\underset{\sim}{X}_{ij}$ is the (1 x p) observation vector for the j^{th} student in the i^{th} group,

and $\overline{\underset{\sim}{X}}_i$ is the (1 x p) mean vector for the i^{th} group.

If one changes the membership constitution of the groups, say by transferring a student from one group to another, or by eliminating a student

from membership in any group, or by changing the number of groups, the W and B matrices will change. Hence, various functions of these matrices might be considered as criteria for objective cluster solutions.

The first function considered is minimum Trace W. Since the diagonal elements of the W matrix are simply sums of squared deviations, minimizing the trace of the W matrix is the same as minimizing the sum of squared Euclidian distances from the data points to their group centroids. This criteria for clustering was first suggested by Sebestyen (1967) and has been used by a number of investigators (Ball and Hall, 1967; MacQueen, 1967; Kendall, 1969). MacQueen labeled cluster solutions using the minimum Trace W criterion "minimum variance partitions."

The second function considered is minimum Determinant W. This function was suggested by Friedman and Rubin (1967); and it follows from consideration of the Wilks' lambda statistic in multivariate analysis. Wilks' lambda statistic

$$\Lambda = |W| \; / \; |T|$$

is used for testing for differences among groups when more than one variable is involved. The magnitude of the differences among groups is inversely related to Λ : i.e., the smaller the Λ, the larger the differences. Since minimum Λ = minimum $|W| \; / \; |T|$ and $|T|$ is a constant, minimum Λ = minimum $|W|$. Hence, minimizing $|W|$ leads to maximal differences among groups as determined by Wilks' lambda statistic.

The third function considered is the maximum largest root of $|B - \lambda W| = 0$. This function is Roy's largest root statistic in multivariate analysis and, as far as this author knows, has not previously

been suggested for application to the clustering situation. Use of this function should tend to maximize differences among groups along the first dimension of discriminant (W^{-1} B) space.

The fourth function considered is maximum sum of roots of $|B - \lambda W| = 0$. This is Hotelling's Trace statistic in multivariate analysis and has been suggested for the cluster analysis application by Friedman and Rubin (1967).

Method

Finding solutions optimizing the criteria specified above is not a trivial matter. The number of ways of partitioning N objects into g groups is very large (see Fortier and Solomon, 1966); the use of an electronic computer and an iterative algorithm is indicated. Toward this end, a computer program (called MIKCA, for Multivariate Iterative K-means Cluster Analysis) was written (McRae, 1971). This program is now described.

The primary procedure used for optimization is the K-means procedure (MacQueen, 1967). The procedure as outlined by MacQueen allows the number of clusters to increase or decrease; MIKCA does not incorporate this option. A supplementary section of MIKCA employs a more time consuming algorithm similar to the algorithm used by Friedman and Rubin (1967).

Information that must be specified by the user includes, in addition to the number of observations and the number of variables, an estimate for the number of clusters. Initial cluster centers are determined by randomly choosing an observation to serve as the initial center for each cluster. All observations are then assigned to the cluster having the closest cluster

center (closest being defined in terms of Euclidian distance); cluster centers are recomputed after each observation is assigned. The order of consideration of the observations is the order of input. After all observations are assigned, the criterion value is computed. This entire procedure is repeated three times; the solution associated with the best criterion value of the three is chosen as the initial cluster solution.

After an initial solution is found, the iterative K-means procedure begins. Once again each observation is considered in the order of input. The observation is assigned to the cluster having the closest cluster center ("closest" being defined as the distance as computed using one of the distance functions described below). After each observation has been assigned, the cluster centers are recomputed. For any given iteration, after all the observations have been considered, the criterion value is computed; if the criterion value is better than the previous iteration, the entire process is repeated; if the criterion value is the same or worse, the program continues on to the supplementary section, called "individual switches."

The individual switches section assigns observations to clusters based directly on the criterion value (as versus a distance function). In addition, the order of consideration of the observations differs from the K-means section of the program. Briefly, this heuristic begins by considering observations in cluster "one." It considers switching each observation in this cluster to each of the remaining clusters; the switch is made if and only if the criterion value improves when the switch is made. After all observations in cluster "one" are considered, the observations in cluster "two" are considered, and so on.

The individual switches heuristic is intended to be a final sharpening process. If any switches are made, then the heuristic will continue to consider

those clusters affected by the switches until no further switches are made. At this point, the final cluster solution is output.

In the iterative K-means section of the program, observations are assigned to clusters based on a distance function specified by the user. One of three distance functions may be specified: Euclidian distance, weighted Euclidian distance, and Mahalanobis distance.

Euclidian distance is defined as

$$d^2 = (X_{ij} - \bar{X}_i)\ (X_{ij} - \bar{X}_i)'$$

where X_{ij} is the j^{th} observation vector in the i^{th} cluster and $\bar{X}_i$ is the mean vector for the i^{th} cluster. This distance function does not take into account either the scale of measurement for the variables or the covariation among the variables.

The weighted Euclidian distance function designed for this program attempts to account for scale differences among the variables. It is defined as

$$d_w^2 = (X_{ij} - \bar{X}_i)\ (\text{diag } W)^{-1}\ (X_{ij} - \bar{X}_i)'$$

The diagonal elements of the within-clusters matrix at any given stage in the analysis reflect the differing variation among the variables. Hence, using this distance function is equivalent to computing distances on variables scaled by the within-cluster standard deviations. Insofar as the within-clusters matrix is a good estimate of the "true" structure in the data, this distance function will adjust for differences due to scale of measurement for the variables.

The third distance function is Mahalanobis distance, which is defined as

$$D^2 = (\underset{\sim}{X}_{ij} - \overline{\underset{\sim}{X}}_i)\ W^{-1}\ (\underset{\sim}{X}_{ij} - \overline{\underset{\sim}{X}}_i)'$$

This distance function takes into account both scale of measurement for the variables and covariation among the variables. Using this distance function is equivalent to computing distances on uncorrelated variables with equal variances.

In summary, optimization of the functions described in the first section of this paper is accomplished by a computer program using primarily an iterative K-means algorithm. This algorithm is supplemented by a more brute force algorithm called "individual switches." The program allows the user to specify one of three distance functions to be used in the iterative K-means section; the distance functions available are designed to adjust for scale and covariation of the variables.

Application

To illustrate the above set of cluster analysis procedures, two sets of data were analyzed. These data were drawn from the tryout sample for the Prescriptive Mathematics Inventory (CTB/McGraw-Hill, 1971), a criterion-referenced mathematics test designed to indicate the knowledge and skills of mathematics for fourth through eighth grade students. The two samples of data will first be described, followed by a description of 18 cluster analysis solutions.

The first sample (Sample A) involved 149 students, each with scores on 10 subsets of the items from the Prescriptive Mathematics Inventory. The students were drawn from six classrooms in three grade schools from a large metropolitan area in the southwestern part of the United States. All students were in grade five at the time of testing. The ten scores were obtained as follows: a tryout edition of the PMI, consisting of 241 items and yielding 34 subscores, was administered. The 34 subscores were reduced to ten scores by adding together subscores in such a manner as to yield ten scores representative of the ten major areas for Level B of the final edition of the PMI. The ten scores, the subscores from which they were drawn, and the number of items contributing to each score are given in Table 1.

The second sample (Sample B) involved 142 students, each with scores on 11 subsets of the items from the Prescriptive Mathematics Inventory. These students were drawn from one junior high school in the same metropolitan area as Sample A. All students were in grade seven at the time of testing. The 11 scores were obtained from a separate tryout edition of the PMI, consisting of 234 items and 41 subscores. The 41 subscores were reduced to 11 scores by adding together subscores in such a manner as to yield 11 scores representative of the 11 major areas in Level C of the final edition of the PMI. The 11 scores, the subscores of the tryout edition from which they were drawn, and the number of items contributing to each score are given in Table 2.

The number of combinations of criterion used, distance function used, and number of clusters desired yields a large number of cluster solutions possible for each data set. A complete exploration of the two data sets is

TABLE 1

Scores, Subscores, and Number of Items for Sample A

Score	Contributing Subscores	Number of Items
Sets	Sets	2
Numeration Systems	Place value, roman numerals	11
Addition	Addition of whole numbers, addition of positive fractions, addition of decimal numbers, number line problems	33
Subtraction	Subtraction of whole numbers, subtraction of positive fractions, subtraction of decimal numbers	15
Multiplication	Multiplication of whole numbers, primes and factors, multiplication of positive fractions, multiplication of decimal numbers	30
Division	Division of whole numbers, division of positive fractions, division of decimal numbers, rounded numbers	33
Properties	Properties	18
Mathematical Sentences	Number sequences, missing addends and factors, mathematical sentences	21
Measurement	Denominate numbers, measurement	29
Non-metric Geometry	Geometry	9
TOTAL		201

TABLE 2

Scores, Contributing Subscores, and Number of Items for Sample B

Score	Contributing Subscores	Number of Items
Sets	Sets	10
Numeration Systems	Place value, numerals	4
Operations	Number line problems, positive fractions, negative fractions, rounded numbers, decimal numbers, integers, missing digits, transforms, missing addends and factors	77
Properties	Properties	24
Mathematical Sentences	Number theory, mathematical sentences	14
Non-metric Geometry	Geometry, ratio	22
Percent	Percent	6
Functions and Graphs	Functions and graphs	8
Measurement	Measurement, geometric computations, hour clock, significant digits	14
Statistics & Probability	Statistics, probability	14
Trigonometry	Trigonometry	4
TOTAL		197

not attempted in this paper; rather, only analyses illustrative of the technique and its results are given. First, for Sample A, three solutions representing 3 different pretreatments of the data are given and discussed. Then, for each data set, eight solutions representing the most popular of options (Trace $\underset{\sim}{W}$ with Euclidian distance, Determinant $\underset{\sim}{W}$ with Mahalanobis distance) are presented and discussed.

Before obtaining cluster solutions, a judgement concerning pretreatment of the data must be made. For the data at hand the number of items contributing to each raw score varies widely; hence, some pretreatment is indicated. To illustrate what happens to solutions under various pretreatments, Trace $\underset{\sim}{W}$, Euclidian distance, three cluster solutions were obtained using three types of scores: (1) raw scores (no pretreatment), (2) standardized score (z-scores), and (3) percent scores. The results of these analyses are given in Table 3.

The three solutions are remarkably similar in that they yield low, medium, and high profile clusters (the clusters were permuted for presentation in Table 3, putting the "low" cluster first, the "medium" cluster second, and the "high" cluster last). This pattern of results will recur. A closer look at the results shows a remarkable similarity between the raw score and z-score solutions; in fact, 136 of the 149 observations are assigned to the same cluster by these two solutions. The percent score solution differs somewhat, offering a slightly clearer resolution between the "low" cluster and the "medium" cluster. Based partially on these results, the remaining analyses presented in this paper were done using percent scores.

In obtaining cluster analysis solutions, one generally does not know before the analysis exactly how many clusters best represent his data. It would be nice to have an indication of the best representation; using MIKCA this is

Table 3

TRACE W, EUCLIDIAN DISTANCE THREE CLUSTER SOLUTIONS

FOR THREE PRETREATMENTS OF THE DATA

	RAW SCORES			Z-SCORES			PERCENT SCORES			MAX. RAW
	Low	Med	High	Low	Med	High	Low	Med	High	SCORE
1.	1.365	1.879	1.821	1.382	1.846	1.897	45.6	100.0	90.7	2
2.	1.942	3.414	5.590	1.945	3.600	6.069	20.9	23.0	48.0	11
3.	9.865	14.741	23.974	10.000	15.954	24.690	29.6	40.5	64.8	33
4.	5.058	7.948	10.923	5.127	8.462	10.966	31.8	47.4	68.4	15
5.	4.885	10.155	13.538	5.073	10.369	14.414	18.0	28.6	40.9	30
6.	2.327	6.672	11.615	2.382	7.246	12.379	7.9	15.2	31.8	33
7.	5.096	10.897	13.590	5.2[illegible]	11.169	14.172	29.4	45.1	77.4	18
8.	2.346	4.914	7.462	2.382	5.200	7.897	12.5	18.6	32.6	21
9.	4.346	9.638	15.077	4.691	10.092	15.828	17.5	25.9	47.3	29
10.	1.096	1.810	3.974	1.091	1.892	4.621	11.8	12.9	43.2	9
Size:	52	58	39	55	65	29	34	61	54	

possible. To illustrate how this is done, solutions for 2, 3, 4, and 5 clusters were obtained for Trace W, Euclidian distance, and Determinant W, Mahalanobis distance for each of the two samples. The results of these analyses are summarized in Table 4. Rather than the actual value of Determinant W, the value of Log $|T| / |W|$ is given in accord with the recommendation of Friedman and Rubin (1967).

The results summarized in Table 4 do not strongly indicate which solution is best representative of the data. The Trace W, Euclidian distance solutions show relatively smooth drops in the Trace W values as the number of clusters increases. The Log $|T| / |W|$ values for Sample A do show that not much change occurred between $g = 3$ and $g = 4$, indicating that the 3-cluster solution is about as efficient as the 4-cluster solution in describing the data. The Sample B Log $|T| / |W|$ values do not show the same effect.

The lack of indication of which solution best represents the data is better understood by considering the 2 and 3 cluster solutions for Sample A, Trace W, Euclidian distance. The cluster centroids for each cluster are plotted in Figures 1 and 2. As is easily seen, the solutions are essentially unidimensional: i.e., a cluster high on one variable tends to be high on all variables, a cluster low on one variable tends to be low on all variables, and so on. All solutions obtained with the IMI data tended to show this type of pattern. The recommendation coming from these considerations would be, then, to determine the number of clusters desired on grounds other than the trend of Trace W or Log $|T| / |W|$ values. Since the results are essentially univariate, one would do about as well to sum the ten variables and cluster the students based on the total score.

TABLE 4

Summary of Results for Sample A and Sample B
g = 2, 3, 4, 5 for Trace $\underset{\sim}{W}$
Euclidian distance and Determinant $\underset{\sim}{W}$, Mahalanobis distance

Sample A

Number of Clusters	Trace $\underset{\sim}{W}$	Log $\lvert\underset{\sim}{T}\rvert$ / $\lvert\underset{\sim}{W}\rvert$
2	41.52	0.684
3	34.35	1.402
4	29.50	1.512
5	26.91	1.991

Sample B

Number of Clusters	Trace $\underset{\sim}{W}$	Log $\lvert\underset{\sim}{T}\rvert$ / $\lvert\underset{\sim}{W}\rvert$
2	44.53	0.708
3	37.40	1.104
4	33.00	1.388
5	30.43	1.827

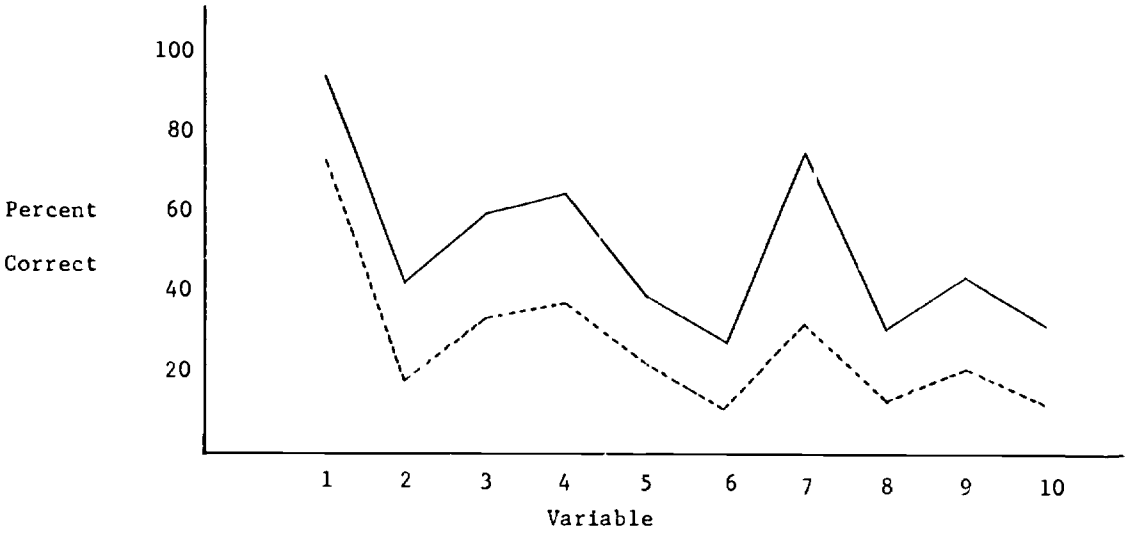

Figure 1

Sample A: Trace W, Euclidian distance, Two cluster solution

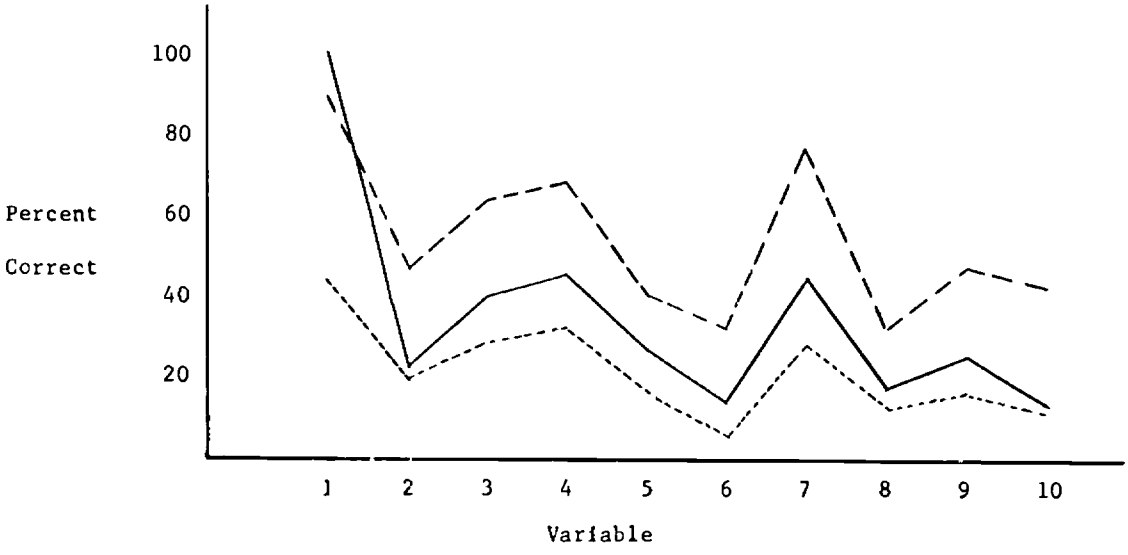

Figure 2

Sample A: Trace W, Euclidian distance, Three cluster solution

Discussion and Conclusion

The set of procedures described in the first part of this paper represent an objective solution to the problem of grouping students. Many questions concerning this technique still must be answered before confident, general use can be made of it. Some of these questions are now discussed.

One obvious question is which clustering criterion yields the best results. There are certain theoretical considerations which favor the criteria based on multivariate analysis; primary among these considerations is the fact that the criteria based on multivariate analysis use the entire $\underset{\sim}{B}$ and $\underset{\sim}{W}$ matrices rather than just the diagonal elements. This means that covariation among the variables enters into the clustering solutions. In addition, the use of the Mahalanobis distance function with the multivariate analysis criteria "equates" the variables for scale and covariation during the solution process. Among the multivariate analysis criteria, the largest root criterion is clearly best for finding maximal unidimensional solutions; the criteria based on Wilks' lambda and Hotelling's trace would clearly be superior if more than one dimension is involved.

Empirical results, both on artificial data and on real data, are also needed to ascertain the types of data for which the use of each criterion is warranted. Along these lines, Friedman and Rubin (1967) report that the Hotelling's trace criterion tends to give unidimensional solutions whereas the Determinant $\underset{\sim}{W}$ criterion does not. Hence, from these results, indications are that the Determinant $\underset{\sim}{W}$, Mahalanobis distance solutions may be the best of the multivariate analysis type solutions.

Another question which must be answered is the question of efficiency and optimality in the computer algorithm. The algorithm described above has some undesirable features, most notably the manner in which the initial cluster solution is obtained. Due to the random element in the initial cluster part of the algorithm, differing solutions can be obtained with differing orders of input. This situation can be used to advantage to obtain an indication of "strength of clustering"; this could be done by re-running the data under a variety of input orders, using the stability of cluster results to indicate "strength of clustering." ("Strength of clustering" is a vague term; what is meant is the general notion of whether the clusters obtained are significant and replicable as versus random artifacts of the forced partitioning.) A preferable solution to the initial clustering problem would be to fix the order of consideration of the observations; the trick here is to find a rule for fixing the order that yields "optimal" results for a variety of data types. Research effort along these lines is continuing.

Another problem that has surfaced with the use of this technique is that the technique tends to find clusters of roughly equal size. Scott and Symons (1970) report that if clusters are of disparate size, for instance if one cluster has five times as many elements as another, the technique tends not to be able to arrive at the appropriate solution. To remedy this, they suggest another criterion: one based on individual within-cluster determinants. They suggest minimizing

$$\prod_{i=1}^{g} |W_i|^{n_i}$$

where $$W_i = \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)' (X_{ij} - \bar{X}_i).$$

This criterion fits well with a modification of the Mahalanobis distance function suggested by Chernoff (1970) for the clustering situation. This modification defines Mahalanobis distance as

$$D^2 = (\underset{\sim}{X}_{ij} - \overline{\underset{\sim}{X}}_i)\ \underset{\sim}{W}_i^{-1}\ (\underset{\sim}{X}_{ij} - \overline{\underset{\sim}{X}}_i)'$$

where $\underset{\sim}{W}_i$ is defined as above. To date, no empirical work has been done either with the criterion suggested by Scott and Symons or with the Mahalanobis distance function suggested by Chernoff.

Finally, there are a number of things that can be done to extend the technique. One of the things would be to allow for a weighting of the variables as specified by the user. The user may want a solution that, on theoretical grounds, weights one score twice as heavily as another score. Another extension of the technique would be to allow analysis on a reduced set of variables, for instance by analyzing a set of r principal component scores derived from the p x p correlation matrix. Since the number of variables is a very important determinant of the computer time required for solution, incorporating this option could prove to be quite time saving. It would also be nice to provide graphic output of the results; the best way to do this seems to be to plot the scores in the first two dimensions of discriminant ($\underset{\sim}{W}^{-1}\underset{\sim}{B}$) space. Research effort on incorporating these options into the procedure is continuing.

In summary, then, this paper describes a cluster analysis technique that allows for completely objective grouping. The options open to the user are described and discussed. Solutions illustrative of the technique using data from the Prescriptive Mathematics Inventory are given. The general

conclusion of this paper is that although much work still needs to be done, the technique represents a promising method for objectively grouping students to minimize classroom heterogeneity.

References

1. Ball, G. H. and Hall, D. J. A clustering technique for summarizing multivariate data. Behavioral Science, Vol. 12 (2), March, 1967.

2. Chernoff, H. Metric considerations in the K-means method of cluster analysis. Paper given at the First Annual Meeting of the Classification Society, Columbia, Ohio, April, 1970.

3. CTB/McGraw-Hill. The Prescriptive Mathematics Inventory. 1971.

4. Fortier, J. J. and Solomon, H. Clustering Procedures. In International Symposium on Multivariate Analysis, edited by P. R. Krishnaiah, Academic Press, 1966, pp. 493-506.

5. Friedman, H. P. and Rubin, J. On some invariant criteria for grouping data. J.A.S.A., (62), 320, pp. 1159-1178.

6. Kendall, M. G. Cluster Analysis. Scientific Control Systems Ltd. London, England, July, 1968.

7. MacQueen, J. Some methods for classification and analysis of multivariate observations. In The Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, U. of California Press, 1967.

8. McRae, D. J. MIKCA: A FORTRAN IV Iterative K-means Cluster Analysis Program, Behavioral Science, in press.

9. Sebestyen, G. S. Decision-making Processes in Pattern Recognition. ACM Monograph Series, Macmillan, New York, 1962.

10. Scott, A. J. and Symons, M. J. Clustering methods based on likelihood ratio criteria. Institute of Statistics Mimeo Series No. 710, University of North Carolina, Chapel Hill, North Carolina, September, 1970.